Information about the data set

EEG and behavioral data of thirteen people were recorded by members of the Neurotechnology Group at Technische Universität Berlin. Details of the study are published in

Wenzel M A, Almeida I and Blankertz B. Is Neural Activity Detected by ERP-based Brain-Computer Interfaces Task Specific? *PLOS ONE*. 2016. To appear.

We kindly ask you to cite this paper if you use the data. The participants of the study gave their informed written consent (a) to take part in the experiment and (b) to the publication of the data in anonymized form without personal information. The study was approved by the ethics committee of the Department of Psychology and Ergonomics of the Technische Universität Berlin (reference BL 02 20140520).

How to load the data

The **EEG data** are stored in the file "EEG.h5" as Pandas DataFrame in the HDF5 format (extract the zipped file first). The data can be loaded with the following Python commands, given that Python (https://www.python.org/, tested with version 3.5.2) and the Pandas package (http://pandas.pydata.org/, tested with version 0.19.0) are installed. The free "Anaconda" is recommended (http://continuum.io/downloads) for an easy installation procedure of all packages and dependencies. Cleaned EEG data, where artefacts have been rejected, can be obtained from the file "EEG_artefacts_rejected.h5" in the exact same manner.

```
import os
import pandas as pd
YOUR_DIRECTORY = '/home/username/data/' # adapt to your setting
fname = os.path.join(YOUR_DIRECTORY,'EEG.h5') # load either 'EEG.h5' or
'EEG_artefacts_rejected.h5'
df = pd.read_hdf(fname, 'df')
```

The two dimensional DataFrame "df" contains the EEG signal at all channels (columns) and time points (rows). The continuous multichannel EEG time series have been preprocessed (as detailed in the paper) and segmented in stimulus-aligned epochs (time = 0 ms corresponds to the stimulus onset). The EEG channel names can be listed by entering "`df.columns`" in the Python shell. Note that the first column "target" indicates if the participants looked at a target stimulus (1) or not (0) at that time.

The hierarchical index of the DataFrame (enter "`df.index.names`") informs for each row about the respective experimental condition, the participant, the EEG epoch and the time in relation to the stimulus onset. The first/last five rows of the (very large) DataFrame can be printed by entering "`df.head()`" or "`df.tail()`"

Pandas allows various operations on the data in few lines of code such as averaging over all EEG epochs of all participants, when they looked at a target stimulus. Each experimental condition, each time point and each channel is treated separately.

```
ga_target = df[df['target']==1].mean(level=['condition','time'])
```

The "target" column and the electrooculogram channel can be removed with

```
ga_target = ga_target.drop(['target','EOG'], axis=1)
```

The experimental condition "A" can be selected with

```
ga_target.loc['A']
```

The **behavioral data** for the three experimental conditions C, A and M are stored in "behavior.zip" as zipped archive of files in the "*.npy" format (there is no need to extract the zipped file). You can load the data with the following Python commands. You may need to additionally install the Python package NumPy (http://www.numpy.org/, tested with version 1.11.2), which also comes with "Anaconda".

```
import numpy as np
import os
YOUR_DIRECTORY = '/home/username/data/' # adapt to your setting
fname = os.path.join(YOUR_DIRECTORY,'behavior.npz')
files = np.load(fname)
C = files['C']
A = files['A']
M = files['M']
```

The correct answer and the answer entered by each participant in each task repetition can be accessed like this:

```
p = 0 # Participant nr. 1 of 13. Indexing starts with zero in Python.
r = 17 # Task repetition nr. 18 of 20.

C[p][r][0] # Correct answer in condition C
C[p][r][1] # Entered answer in condition C

A[p][r][0] # Correct answer in condition A
A[p][r][1] # Entered answer in condition A

M[p][r][0] # Correct coordinates in condition M
M[p][r][1] # Entered coordinates in condition M
```